

# Uniform generation of random acyclic digraphs

Jack Kuipers<sup>a</sup>, Giusi Moffa<sup>b</sup>

<sup>a</sup>*Institut für theoretische Physik, Universität Regensburg, D-93040 Regensburg, Germany*

<sup>b</sup>*Institut für funktionelle Genomik, Universität Regensburg, Josef Engertstraße 9, 93053 Regensburg, Germany*

---

## Abstract

We show how to sample acyclic digraphs uniformly at random through recursive enumeration. This provides an exact method which avoids the convergence issues of the alternative Markov chain methods. The limiting behaviour of the distribution of acyclic digraphs also allows us to sample arbitrarily large acyclic digraphs. Finally we discuss how to include various restrictions in the combinatorial enumeration for efficient uniform sampling of the corresponding graphs.

**Keywords:** Random graph generation, acyclic digraphs, recursive enumeration, Bayesian networks

---

## 1. Introduction

Directed acyclic graphs (DAGs) are the basic representation of the structure underlying Bayesian networks, which in turn represent multivariate probability distributions (Lauritzen, 1996; Neapolitan, 2004). They are largely used in many fields of applied statistics with especially important applications in biostatistics, such as the learning of epistatic relationships (Jiang et al., 2011). The estimation of DAGs or their equivalence class is a hard problem and methods for their efficient reconstruction from data is a very active field of research: a recent review is given by Daly et al., 2011 while some new methodological developments for estimating high dimensional sparse DAGs are discussed by Kalisch and Bühlmann, 2007; Colombo et al., 2012. For simulation studies aimed at assessing the performance of learning algorithms which reconstruct a graph from data, it is crucial to be able to generate uniform samples from the space of DAGs so that any structure related bias is removed. The only currently available method relies on the construction of a Markov chain whose properties ensure that the limiting distribution is uniform over all DAGs with a given number of vertices  $n$ . The strategy is based on a well known idea first suggested by Madigan and York (1995) as a Markov Chain Monte Carlo (MCMC) scheme in the context of Bayesian graphical models to sample from the posterior distribution of graphs conditional on the data. A specific algorithm for uniform sampling of DAGs was first provided by Melançon et al. (2001), with the advantage over the standard MCMC scheme of not requiring the evaluation of the sampled graphs' neighbourhood size, at the expense of a slower convergence. The method was later extended by Ide and Cozman (2002); Ide et al. (2004); Melançon and Philippe (2004) to limit the sampling to restricted sets of DAGs. An R implementation was also recently provided by Scutari (2010). Since Markov chain based algorithms pose non-negligible convergence and computational issues, in practice random upper or lower triangular adjacency matrices are often sampled to generate random ensembles for simulation studies [as for example implemented in the *pcalg* R package of Kalisch et al. (2012)]. This method however does not provide uniformly distributed graphs on the space of DAGs and could for example perform poorly to obtain starting points for hill-climbing algorithms or slowly converging Markov chains by increasing the risk of remaining within a small neighbourhood of certain graphs and more inefficiently exploring the space. Likewise uniform sampling allows the correct evaluation of reconstructing algorithms. Finally, when evaluating the prevalence of certain features in a population, a uniform sample is essential. Here we therefore present a sampling strategy based on the recursive enumeration of DAGs but where no explicit listing is required.

---

*Email addresses:* jack.kuipers@physik.uni-regensburg.de (Jack Kuipers), giusi.moffa@klinik.uni-regensburg.de (Giusi Moffa)

## 2. Acyclic digraphs

A directed graph or digraph on  $n$  vertices has up to one directed edge between each pair of vertices, from node  $i$  to node  $j$  with  $1 \leq i, j \leq n$ . Acyclic digraphs or DAGs are those which admit no cycles so that there are no paths along the directed edges from a node to itself. The total number  $a_n$  of labelled DAGs with  $n$  nodes, which is sequence A003024 in Sloane (2012), and its asymptotic behaviour are respectively (Robinson, 1970, 1973; Stanley, 1973)

$$a_n = \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} 2^{k(n-k)} a_{n-k}, \quad a_n \sim \frac{n! 2^L}{M q^n}, \quad (1)$$

with  $L = \binom{n}{2} = n(n-1)/2$  and where  $M = 0.574 \dots$  and  $q = 1.48 \dots$  are constants. The number of unlabelled DAGs was found later by Robinson (1977) and is sequence A003087 in Sloane (2012).

The directed edges of a DAG can be represented as entries with value 1 of a vertex adjacency matrix with remaining entries 0. Being acyclic the matrix can be made upper triangular by relabelling the nodes, which lies behind the proof of McKay et al. (2004) that the number of  $n \times n$  (0,1)-matrices with only positive real eigenvalues is also  $a_n$ . Hence DAGs can be easily generated by sampling each of the  $L$  elements of an upper triangular matrix from a Bernoulli with probability  $p = 1/2$  and then permuting the  $n$  labels, but they are non-uniform since several permutations may correspond to the same graph. For example, the DAG with no arcs (the 0 matrix) would be counted  $n!$  times, while those with the most arcs (when all  $L$  upper triangular elements are 1) only once. Choosing  $p \neq 1/2$  may reduce the overall non-uniformity but because the overcounting is sensitive to the structure of the DAG, not remove it entirely.

### 2.1. Markov chain method

The algorithm of Melançon et al. (2001) starts with a given DAG, with no arcs say, and proceeds by repeatedly sampling an ordered pair of vertices from the  $n$  available. If an arc joins the pair in that direction it is deleted, otherwise it is added as long as the graph remains acyclic. Addition and deletion of an arc are inverse operations so that the resulting transition matrix  $T$  is symmetric. As adding an arc to any pair  $(i, i)$  would immediately create a cycle the probability to stay with the same graph is nonzero and at least  $1/n$ . One could exclude all but one such pair from the sampling to reduce this probability and marginally improve the convergence of the chain. Finally a path exists between any pair of DAGs since deleting all the arcs in turn leads to the empty DAG and the steps are reversible. These three properties ensure that the stationary distribution of the Markov chain is uniform over the space of graphs.

The Markov chain method is very elegant, as no particular knowledge of DAGs is required apart from checking whether the graphs remain acyclic, which can be done in a time of the order between  $n$  and the number of edges (typically around  $n^2$ ) and on average in a time of order  $n \log(n)$  (Alon and Rodeh, 1978; Melançon et al., 2001). The issue is how quickly the chain converges. The longest path is between a DAG with the most arcs ( $L$ ) and the same graph with all the edges reversed. Therefore after  $2L$  steps we obtain an irreducible transition matrix  $U = T^{2L}$ , whose elements though are far from uniform. First removing all the arcs in  $L$  steps has a transition probability of  $L!/n^{2L}$ , and the same adding them back due to the symmetry. The  $L$  additions and deletions can be ordered in  $\binom{2L-2}{L-1}$  ways, though the probability of different orderings is lower. This leads to a minimum transition probability  $< (2L)!/n^{4L}$  which is actually smaller than  $1/a_n^2$ , much less than the required uniform value of  $1/a_n$ .

The difference between the maximum and minimum of the elements of  $U^k$  decays at least exponentially  $\sim \exp(-\alpha k)$  with a rate  $\alpha$  depending on the largest eigenvalue of  $U$  below that at 1. The probability for any DAG with the most arcs to remain unchanged is  $1/2 + 1/2n$  so the maximum element of  $U^{1/2}$  is at least  $1/2^L$ . To approach the uniform distribution it should be reduced by a factor of around  $n!/q^n$ , in line with (1), suggesting that  $k$  is at least of order  $n \log(n)/\alpha$ . Since the largest element of  $U$  is certainly  $< 1$  the order of  $k$  is at most  $n^2/\alpha$ . Now what matters for the speed of convergence is how  $\alpha$  depends on  $n$ . Proving lower bounds is extremely hard and a generally known one, given by twice the minimum entry of  $U$ , is not enough to ensure that the algorithm is useful in practice since it would require  $k$  much larger than  $a_n$  for convergence. For DAGs with 2 or 3 nodes  $T$  can be easily filled out and we find  $\alpha = 0.575$  and  $\alpha = 0.594$  respectively, and with a reduced probability of not moving,  $\alpha = 0.811$  and  $\alpha = 0.776$  instead. Its behaviour should be verified, but as long as  $\alpha$  does not decrease asymptotically with  $n$ , a uniformly sampled DAG can be obtained by throwing away the first  $O(n^4)$  steps of the Markov chain. Analogously  $O(n^4)$  steps should also be discarded between further samples to ensure they are independent. These conditions should be observed when using implementations like that in Scutari (2010), with the entire algorithm then becoming  $O(n^5 \log(n))$  and limiting the size of graphs that can be sampled.

### 3. Enumeration method

In order to obtain a uniform sample directly, we return to the enumeration of labelled DAGs detailed in Robinson (1970, 1977). Nodes with no incoming arcs are called ‘outpoints’ and can be used to further classify and recursively enumerate DAGs, which have at least one since they are acyclic. Let  $a_{n,k}$  be the number of labelled DAGs with  $n$  nodes of which  $k$  are outpoints ( $1 \leq k \leq n$ ). Removing the latter and the arcs originating from them leaves a smaller DAG with  $m = n - k$  nodes, with  $s$  outpoints say. Reversing the process, the required DAGs can be built by adding  $k$  new outpoints and allowing all the possible connections from them to the previous nodes. For each of the  $m - s$  non-outpoint nodes there are  $2^k$  possibilities of having an arc or not, while each of the  $s$  old outpoints must be connected to at least once, giving a factor of  $2^k - 1$ . Finally the labels can be rearranged in  $\binom{n}{k}$  ways, giving the recursions (Robinson, 1970, 1977)

$$a_{n,k} = \binom{n}{k} b_{n,k}, \quad b_{n,k} = \sum_{s=1}^m (2^k - 1)^s 2^{k(m-s)} a_{m,s} \quad (2)$$

with  $a_{n,n} = 1$  the DAGs with no arcs and  $b_{n,k}$  auxiliary quantities useful later. The total number of DAGs with  $n$  nodes is  $a_n = \sum_{k=1}^n a_{n,k}$  and can be turned into the more compact form in (1) by means of the inclusion-exclusion principle. The separation in terms of outpoints however is the key for our strategy to sample labelled DAGs uniformly.

#### 3.1. Recursively generating the outpoints

The first step in our algorithm is to compute all the integers  $a_{n,k}$  (and  $b_{n,k}$ ) as well as the totals  $a_n$ . From (1) it follows that the number of bits needed to store them grows like  $L$ , while  $2n + 1$  integers are needed for each  $n$  (since  $1 \leq k \leq n$ ). Finally each of them is obtained by adding an order of  $n$  smaller integers (multiplied by the binomial function) so that the process becomes  $O(n^4)$ . However, they only need to be calculated once and then stored for sampling arbitrarily many DAGs. Later we also discuss how to avoid their explicit evaluation.

The second step is, given  $n$ , to sample the number of outpoints. An integer is drawn between 1 and  $a_n$ , for example by sampling from a Bernoulli( $\frac{1}{2}$ ) the digits of a binary string long enough to represent  $a_n$  (if the resulting number is larger than  $a_n$  the string is rejected and resampled, which happens with probability less than a half). The chosen number of outpoints is the minimum  $k$  for which the partial sum  $\sum_{i=1}^k a_{n,i}$  is larger than the sampled integer.

The final DAG will be obtained by connecting the  $k$  outpoints to a smaller one of size  $m$ , whose number of outpoints can be taken to be the minimum  $s$  for which the partial sum  $\sum_{i=1}^s (2^k - 1)^i 2^{k(m-i)} a_{m,i}$  is larger than a randomly drawn integer between 1 and  $b_{n,k}$ . Similarly a DAG with  $m$  nodes and  $s$  outpoints can be sampled by drawing an integer between 1 and  $b_{m,s}$ . The number of outpoints  $k_i$  removed at each step can so be recursively sampled until their  $\sum_i k_i = n$ . In total we need to sample and manipulate  $I \leq n$  integers whose size in bits is at most of order  $L$ , so that generating the sequence is  $O(n^3)$ . Though we proposed sampling and integer at each step, the original one between 1 and  $a_n$  could be equally reused to obtain the entire sequence of outpoints.

#### 3.2. Reconstructing the DAG

To represent the sampled DAG we fill a null  $n \times n$  adjacency matrix from the top left to obtain a lower triangular matrix. Since the recursion ends when all the remaining nodes are outpoints, the first  $k_I \times k_I$  elements stay zero. These must connect to at least one of  $k_{I-1}$  newly added outpoints which must not connect to each other, hence the elements of each column from 1 to  $k_I$  in rows  $k_I + 1$  to  $k_I + k_{I-1}$  can be sampled from a Bernoulli( $\frac{1}{2}$ ). Any samples where all the elements in a column are 0 can be rejected columnwise (alternatively a direct binomial sampling with this case excluded can be performed).

When adding an additional  $k_{I-2}$  nodes, the elements in columns 1 to  $k_I$  and rows  $k_I + k_{I-1} + 1$  to  $k_I + k_{I-1} + k_{I-2}$  are again sampled from a Bernoulli( $\frac{1}{2}$ ), while for each column  $k_I + 1$  to  $k_I + k_{I-1}$  we need to ensure that not all new entries are 0. Completing the adjacency matrix is a step of order  $n^2$  since up to  $L$  integers need to be sampled. The final sampled DAG can be obtained by permuting the labels only at the end, rather than at each step as suggested by (2), since each permutation must be equally likely. Although different permutations may still correspond to the same DAG, drawing the number of outpoints at each step is weighted so that the resulting sample is perfectly uniform.

Table 1: The relative occurrence of the number of outpoints in large DAGs,  $A_k$  multiplied by  $10^{10}$ .  $A_8$  is approximately  $2.2 \times 10^{-12}$ , so  $k > 7$  can be excluded at this level of accuracy.

|       |               |       |             |       |         |       |    |
|-------|---------------|-------|-------------|-------|---------|-------|----|
| $A_1$ | 5 743 623 733 | $A_3$ | 564 645 435 | $A_5$ | 566 517 | $A_7$ | 15 |
| $A_2$ | 3 662 136 732 | $A_4$ | 29 023 072  | $A_6$ | 4 496   | $A_8$ | 0  |

### 3.3. Arbitrarily large DAGs

The number of outpoints sampled at each step is small and most often just 1. In fact the fraction of DAGs with  $k$  outpoints among  $n$ ,  $A_k = \lim_{n \rightarrow \infty} a_{n,k}/a_n$ , converges to within  $10^{-10}$  by  $n = 20$ . For  $n > 20$ , within the uniformity limits of 32 bit integer sampler,  $k$  could be directly sampled from the limiting distribution in Table 1. Given the initial  $k$  the next value  $s$  in the sequence of outpoints can be drawn from the limiting conditional probabilities

$$B_{s|k} = \left(1 - \frac{1}{2^k}\right)^s \frac{A_s}{Z_k}, \quad Z_k = \sum_s \left(1 - \frac{1}{2^k}\right)^s A_s \quad (3)$$

which follow from (2). The  $k$ -dependent weight further shifts the distribution towards low  $s$  so it would be computationally meaningful to store the small number of relevant probabilities. The procedure is iterated until we are left with 20 nodes or fewer, when the tabulated distributions are no longer accurate enough. For  $n \leq 20$  one could tabulate the exact distributions or simply return to the enumeration method. Generating the sequence of outpoints is now of order  $n$  so that a uniform (at scales above  $10^{-10}$ ) DAG can be sampled in  $O(n^2)$  allowing one to move to very large graphs.

The accuracy of the convergence to the limiting distributions of  $A_k$  nearly exactly doubles when doubling  $n$ . To obtain uniformity on the scale of a 64 bit integer sampler we can use (a more accurately calculated distribution of)  $A_k$  for  $n > 40$  and return to the exact enumeration for  $n \leq 40$ .

## 4. Restrictions and variations

A promising application of the Markov chain method, as particularly pursued in Ide and Cozman (2002); Ide et al. (2004), is to impose restrictions on DAGs sampled. For example the number of edges and the degrees of the nodes can be restricted by simply rejecting any steps producing graphs which do not meet the requirements. More complicated Markov chains are needed for sampling polytrees and DAGs with restricted density (Ide and Cozman, 2002; Ide et al., 2004) but the speed of convergence to uniformity remains an issue. Although the space of restricted graphs is smaller, the probability of not moving for graphs on the border of the restricted set increases due to the additional rejections. If no recursive enumeration is possible for the restricted graphs, however, the Markov chain approach remains the only one available, but for restrictions which can be incorporated combinatorially, direct enumeration again provides a more efficient method and we discuss some examples below.

### 4.1. Connected DAGs

The main restriction imposed in Ide and Cozman (2002); Melançon and Philippe (2004) is that DAGs be weakly connected (admitting a path between every pair of nodes in the underlying undirected graph). They were also counted by Robinson (1973) and are sequence A082402 in Sloane (2012). The Markov chain method can be simply modified so to only delete an arc if the resulting graph is also connected (Ide and Cozman, 2002) or, more efficiently, instead to reverse its direction otherwise (Melançon and Philippe, 2004). Checking for connectedness is of order of the number of arcs making the algorithm  $O(n^6)$  and certainly slower since at each step either connectedness or acyclicity of the proposed graph needs to be checked. The fraction of connected DAGs tends to 1 as  $n$  increases (Robinson, 1973; Bender et al., 1986; Bender and Robinson, 1988); for  $n = 4$  it is 82% (for smaller  $n$  we can enumerate the possibilities by hand), while it is  $> 99\%$  and growing for  $n > 8$ . Therefore the above modifications of the Markov chain, though theoretically interesting, are of no practical use, as it is much more efficient to sample a DAG at random and reject the disconnected ones.

When removing the  $k$  outpoints of a connected DAG the rest breaks up into a set of smaller connected DAGs, hence a recursive combinatorial enumeration (involving the sum over the partitions of  $m$ ) is possible. Again the increase in algorithmic complexity makes a simple rejection step more efficient, apart from possibly for small  $n$ .

#### 4.2. Restricted new connections

Because new outpoints are added at each stage of the recursive enumeration of DAGs, restrictions on the number of incoming arcs the previous nodes can receive from them are the simplest to treat. Allowing up to a maximum  $K$  new connections ( $K \geq 1$  avoids only generating DAGs with no arcs) the recursions change to

$$\tilde{a}_{n,k} = \binom{n}{k} \sum_{s=1}^m \left[ \sum_{i=1}^{\min(k,K)} \binom{k}{i} \right]^s \left[ \sum_{i=0}^{\min(k,K)} \binom{k}{i} \right]^{m-s} \tilde{a}_{m,s} \quad (4)$$

For  $K > k$  the above reduces to (2) while a restriction on the minimum number of new connections can be incorporated analogously. Such restricted DAGs can easily be counted and then sampled as in Section 3. In reconstructing the DAGs new arcs can be uniformly sampled from the binomial distributions in (4). Retaining the limit on the arcs received by the previous outpoints, it is also straightforward to separately limit the connections each new one can send to the previous non-outpoints to a maximum of  $K_n$ . The corresponding DAGs can be sampled through

$$\tilde{a}_{n,k} = \binom{n}{k} \sum_{s=1}^m \left[ \sum_{i=1}^{\min(k,K)} \binom{k}{i} \right]^s \left[ \sum_{i=0}^{\min(m-s,K_n)} \binom{m-s}{i} \right]^k \tilde{a}_{m,s} \quad (5)$$

#### 4.3. Restricted number of children

Because they must be connected to, including previous outpoints when limiting the outgoing arcs from each new one is more complicated. Since arcs only originate from newly added outpoints, this means limiting the children to a certain  $K$ . An expression for the ways  $C(k, m, s, K)$  of linking  $k$  new outpoints may be found by subtracting from the ways of connecting  $k$  nodes to  $m$  (with up to  $K$  arcs each) those which leave any of the  $s$  outpoints unconnected to

$$C(k, m, s, K) = \left[ \sum_{i=0}^{\min(m,K)} \binom{m}{i} \right]^k - \sum_{i=1}^s \binom{s}{i} C(k, m-i, s-i, K), \quad \tilde{a}_{n,k} = \binom{n}{k} \sum_{s=1}^m C(k, m, s, K) \tilde{a}_{m,s} \quad (6)$$

The graph is reconstructed by drawing from the  $C(k, m, s, K)$  possibilities at each step, and formula (6) simply suggests rejecting the cases which leave some of the  $s$  outpoints with no incoming link. It is however more efficient to first obtain the distribution of the number of arcs linking each new outpoint to the old ones when excluding the cases with a total  $< s$ . Samples from it where any old outpoints do not receive arcs are then rejected. Arcs to the remaining nodes can be drawn subsequently. For  $C(k, m, s, K)$  much smaller than the first term on the right of (6), or when the average number of new arcs linking to the old outpoints falls, the acceptance ratio also drops. In this limit it may be convenient to first draw exactly one link to each of them and then sample the remaining arcs among all the nodes, including the  $s$  old outpoints. The configurations where each receives  $l_i$  incoming arcs are overcounted by a factor  $F = \prod_{i=1}^s l_i$ , so the sample should only be accepted with probability  $1/F$ . A balance of the two approaches may help minimise the rejection probability, but this example highlights that just knowing the number of possibilities is not sufficient for reconstruction, and a constructive method to generate them is preferable. A uniformly drawn DAG with a limit on the number of parents can be simply obtained by inverting all the arcs of a DAG with a limited number of children.

#### 4.4. Weighted sample

When generating DAGs by sampling upper triangular matrices it is easy to influence their sparsity by sampling the entries from a Bernoulli with  $p \neq 1/2$ , this weighting however acts on an underlying non-uniform space of DAGs. In practice though it would be desirable to obtain a weighted sample from an underlying uniform distribution, which is achieved by making the probability of a DAG with  $l$  arcs proportional to  $p^l(1-p)^{L-l}$  and including analogous factors in the recursive enumeration. Each of the  $m-s$  non-outpoints may receive no links from the  $k$  new ones with weight  $(1-p)^k$ , one link with weight  $kp(1-p)^{k-1}$  and so on to include all the terms in the binomial expansion of  $[(1-p) + p]^k = 1$ . The possibility that no arcs link to the  $s$  old outpoints is excluded, while no arcs are allowed between the  $k$  new ones out of a theoretical  $\binom{k}{2}$  ways of joining them, leading to the recursion

$$\hat{a}_{n,k} = \binom{n}{k} (1-p)^{\binom{k}{2}} \sum_{s=1}^m (1-(1-p)^k)^s \hat{a}_{m,s} \quad (7)$$



where now  $\hat{a}_{n,n} = (1 - p)^L$  correspond to the DAGs with no arcs out of  $L$  possibilities. We take  $0^0 = 1$  so that  $\hat{a}_{1,1} = 1$  even for  $p = 1$  where only one new outpoint is added at each step. Weighted DAGs can be generated as in Section 3 where links are now added at each step with probability  $p$ . The terms in the recursions are no longer integers in general (though for rational  $p$  the numerator and denominator could be stored as a pair of integers) and exact sampling can be obtained up to the precision limits of the computational implementation.

## 5. Conclusions

The recursive method for generating DAGs we presented has the advantage over the highly adopted strategy of sampling triangular adjacency matrices of producing a uniformly distributed sample and over Markov chain based algorithms of avoiding convergence problems and being computationally more efficient. Moreover the ideas discussed here can also be combined with an MCMC scheme in the inference of Bayesian graphical models. A procedure can be built following the combinatorial construction where moves are suggested by going back a certain number of steps in the sequence of adding outpoints and resampling just the new connections at that step. As they are sampled uniformly, such moves are reversible. To ensure full space reachability, one could introduce some probability of proposing a move uniformly amongst all DAGs or according to the standard MCMC on network structures (Madigan and York, 1995). The idea is similar in spirit to the sampling in the space of orders of Friedman and Koller (2003) and has therefore the potential of greatly improve the mixing properties of the chain produced, with the advantage of possibly proposing structures which are in some sense close in the space of graphs. Since neither the original MCMC on network structures nor the order version appear to be completely satisfactory in practice (Grzegorzcyk and Husmeier, 2008) the possibility outlined above seems promising.

## References

- Alon, I., Rodeh, M., 1978. Finding a minimum circuit in a graph. *SIAM Journal on Computing* 7, 413–423.
- Bender, E.A., Richmond, L.B., Robinson, R.W., Wormald, N.C., 1986. The asymptotic number of acyclic digraphs. I. *Combinatorica* 6, 15–22.
- Bender, E.A., Robinson, R.W., 1988. The asymptotic number of acyclic digraphs. II. *Journal of Combinatorial Theory, Series B* 44, 363–369.
- Colombo, D., Maathuis, M.H., Kalisch, M., Richardson, T.S., 2012. Learning high-dimensional directed acyclic graphs with latent and selection variables. *Annals of Statistics*, In press.
- Daly, R., Shen, Q., Aitken, S., 2011. Learning Bayesian networks: approaches and issues. *The Knowledge Engineering Review* 26, 99–157.
- Friedman, N., Koller, D., 2003. Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks. *Machine Learning* 50, 95–125.
- Grzegorzcyk, M., Husmeier, D., 2008. Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move. *Machine Learning* 71, 265–305.
- Ide, J.S., Cozman, F.G., 2002. Random generation of Bayesian networks, in: *Brazilian Symposium on Artificial Intelligence*, pp. 366–375.
- Ide, J.S., Cozman, F.G., Ramos, F.T., 2004. Generating random Bayesian networks with constraints on induced width, in: *European Conference on Artificial Intelligence*, pp. 323–327.
- Jiang, X., Neapolitan, R., Barnada, M.M., Visweswaran, S., 2011. Learning genetic epistasis using Bayesian network scoring criteria. *BMC Bioinformatics* 12.
- Kalisch, M., Bühlmann, P., 2007. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research* 8, 613–636.
- Kalisch, M., Machler, M., Colombo, D., Maathuis, M.H., Bühlmann, P., 2012. Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software*, In press.
- Lauritzen, S.L., 1996. *Graphical models*. Clarendon Press.
- Madigan, D., York, J., 1995. Bayesian graphical models for discrete data. *International Statistical Review* 63, 215–232.
- McKay, B.D., Oggier, F.O., Royle, G.F., Sloane, N.J.A., Wanless, I.M., Wilf, H.S., 2004. Acyclic digraphs and eigenvalues of  $(0,1)$ -matrices. *Journal of Integer Sequences* 7, 04.3.3.
- Melançon, G., Dutour, I., Bousquet-Mélou, M., 2001. Random generation of directed acyclic graphs. *Electronic Notes in Discrete Mathematics* 10, 202–207.
- Melançon, G., Philippe, F., 2004. Generating connected acyclic digraphs uniformly at random. *Information Processing Letters* 90, 209–213.
- Neapolitan, R.E., 2004. *Learning Bayesian networks*. Pearson Prentice Hall.
- Robinson, R.W., 1970. Enumeration of acyclic digraphs, in: *Proceedings of the second Chapel Hill conference on combinatorial mathematics and its applications*, University of North Carolina, Chapel Hill. pp. 391–399.
- Robinson, R.W., 1973. Counting labeled acyclic digraphs, in: *New directions in the theory of graphs*. Academic Press, New York, pp. 239–273.
- Robinson, R.W., 1977. Counting unlabeled acyclic digraphs, in: *Combinatorial mathematics V. volume 622 of Springer Lecture Notes in Mathematics*, pp. 28–43.
- Scutari, M., 2010. Learning Bayesian networks with the bnlearn R package. *Journal of Statistical Software* 35, 1–22.
- Sloane, N.J.A., 2012. The on-line encyclopedia of integer sequences, published electronically at <http://www.research.att.com/~njas/sequences/>.
- Stanley, R.P., 1973. Acyclic orientations of graphs. *Discrete Mathematics* 5, 171–178.